Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-289-RC2, 2017

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Development of the WRF-CO2 4DVar assimilation system" *by* Tao Zheng et al.

Anonymous Referee #2

Received and published: 11 May 2017

The manuscript, "Development of the WRF-CO2 4DVar assimilation system", describes exactly that. The authors have created linearized model versions of the WRF-Chem CO2 transport mechanisms, validated their performance against finite difference approximations, and demonstrated their utility in a simplified pseudo-data experiment. The introduction covers much of the relevant literature necessary to get to the same starting point as the authors, and we make a few additional suggestions below. The adjoint and tangent linear model developments are described thoroughly, and would be helpful for any person working their way through the code at a later time. The adjoint model evaluation falls a little short, and we provide some suggestions for ways it could be improved. The pseudo-data inversion test, while quite simple and unrealistic, demonstrates that the inversion framework is working. It is a first step that undoubtedly took considerable effort, but needs some improvements in the application of the new

Printer-friendly version

Discussion paper

tool. There is no discussion of the statistical nature of 4D-Var, which is paramount to that method's success with real data and its being labeled a Bayesian inversion technique. We have several specific comments as to how the discussion could be made more precise and also miscellaneous technical corrections.

Major Comments:

- Section 2.3: Incremental 4D-Var is used to optimize nonlinear systems. But the CO2 tracer simulation is inherently linear. Thus, I don't really see at this point what the benefit would be of an incremental formulation, nor how updating the inner loop with an outer loop integration would provide any additional information. Thus, the importance of including both of these methods and their comparisons for CO_2 inversions needs further justification.
- Fig 8: As the authors surely know, adjoint sensitivities should agree with the tangent linear sensitivities to numerical precision. The differences between these sensitivities vs the finite difference sensitivities, given that the latter match the tangent linear sensitivities, is an indication that the adjoint model is not error free. For the cases tested here, the errors are manageable, yet there is no guarantee that the errors would not grow for longer simulations. The authors should thus continue to debug their adjoint code, possibly by performing this type of test around the tangent linear and adjoint code of individual physics components developed here (such as ACM2 PBL mixing). If they can not resolve the code bugs this way, they should at least perform additional tests using different receptor locations and simulations of increasing (and decreasing) length to examine how the numerical errors may be accumulating.
- p. 12, line 7: It is not required that a grid cell be both a receptor and a source to have non-zero sensitivity, as evidenced by Figures 5 and 6. The source and receptor must simply be significant (large source and large perturbation to concentration due to that particular source). Choosing them to be the same grid

cell is likely to produce good agreement between the adjoint and finite difference methods even when the advection, PBL, or convective transport adjoint may be incorrect. Did the authors choose identical grid cells for source and receptor? If so, then additional tests are required to prove that the adjoint code works as described. Additionally, validating the convective transport adjoint and tangent linear codes requires demonstration during a period when the subgrid cumulus parameterization is active and for sources and receptors near that phenomenon. The authors should present some indicator of cumulus activity in an additional figure.

- p. 12, lines 16-17: Similar to the comment above, choosing the observations to be in the lowest model layer reduces the importance of having the adjoint and tangent linear treatments of vertical mixing. The observations should be spread more thoroughly, vertically.

- From a software perspective, I'm a bit confused about the distinction between the present work and that of GH15/16, where the adjoint of the BC tracer is developed. So, essentially the update here is that BC has been changed to CO2, and convective transport has been added? On a broader note, would it be beneficial to the community to view these as two options within a single chemical 4D-Var system, rather than as two different models? I realize this likely results from development of these systems over time, in parallel, but thinking to the future I wonder if a model merge would be in order. To illustrate my point, imagine if rather than a consolidated WRF-Chem model, we had separate WRF-Chem-BC, WRF-Chem-CO2, WRF-Chem-CO, ... etc. models. That would clearly hinder development of the tool as a whole, which shares many common elements across the different tracer simulations.

Further Comments:

1. p. 2, line 33: The reference to Streets 2013 is a bit odd, as that paper is specifically a review of remote sensing based constraints on emissions and focuses mainly on reactive trace gases and aerosols. As the present work doesn't seem geared towards remote sensing observations, some other references to literature on regional CO2 inverse modeling would be a better fit here.

2. p. 2, line 9: Probably more correct to say "instead they directly compute the product of the Jacobian with a forcing vector, which is the gradient used for optimizing the state vector."

3. p. 2, line 10: The notion that posterior error can not be calculated from a variational inversion is outdated. Posterior error can be calculated analytically using the Lanczos vectors from a CG minimization in the incremental 4D-Var framework following Fisher and Courtier (1995) - as currently done in operational weather forecast centers such as ECMWF and the UK Met Office - for minimal additional cost. Efficient posterior error estimate for non-incremental 4D-Var frameworks are described in Bousserez et al., QJRMS, 2015, including previous works on Monte Carlo (e.g., Chevallier et al., JGR, 2007) and stochastic (Rabier and Courtier, QJRMS, 1992) methods.

   Fisher M, Courtier P. 1995. Estimating the Covariance Matrices of Analysis and Forecast Error in Variational Data Assimilation, Technical Memorandum 45. ECMWF: Reading, UK.

4. The introduction states that both bottom-up and top-down approaches are used, but does not say why that is the case. It is recommended to move the first paragraph of Section 4 to the introduction.

5. p. 2., line15: The GEOS-Chem CO_2 4D-Var system is also part of JPL's Carbon Monitoring System, e.g. Liu, Bowman, Lee, et al., Tellus B, 2014; Liu, Bowman, and Lee, JGR, 2016.

Printer-friendly version

Discussion paper

6. p. 2, line 18: Also Chevallier et al., JGR, 2007.

7. p. 2: 2: For regional CO2 inversions, the list isn't entirely complete, see also Alden, Miller, Gatti, et al., Global Change Biology, 2016; Chan, Chan, Ishizawa, et al., GMDD, 2016. There are others, but I think it suffices to say the literature review could be a bit more comprehensive (or, alternatively, scoped / phrased as to be more narrow).

8. p. 2, line 33: Here and in several other places, the authors use the phrase "influence function" without every having defined it.

9. Equations 3,4: define superscript n in this context.

10. Equations 6,7: I understand incremental 4D-Var, but I think still the authors should rigorously define the superscript n in this context for the sake of completeness, which I believe should differentiate between inner and outer loops. Also, incremental 4D-Var is usually employed with a square-root preconditioning, which I don't see here.

11. Fig 1: The way that the observations fit into this diagram doesn't make sense, since currently it implies the arrow coming out of the right side of the Simulated box passes information both to the right and left. Some separate arrows from the Observation box seem to be needed.

12. Fig 2: Despite the caption, this doesn't really show how the CG method is implemented to anyone already not familiar enough with incremental 4D-Var to know that it likes between the "no" and Tangent linear model and involves an updated estimate to the (preconditioned) increment.

13. p. 3, lines 11-13: This sentence is grammatically incorrect, the phrasing is confusing, and the conclusion is drawn weakly. What is the "potential" that online

transport based inversion systems have demonstrated? At a bare minimum, add a reference (e.g., Grell and Baklanov, 2011) or remove that statement.

14. p. 3, line 18, p .6, line 3, and p .6, line 23: The original reference for WRFPLUS is Xiao et al. (2008) [DOI: http://dx.doi.org/10.1175/2008MWR2235.1]. The version you use (v3.6) includes the work by Zhang et al. (2013) [mentioned elsewhere in your text], and should be included in these references. Huang et al. (2009) specifically used WRFPLUS for 4D-Var in WRFDA, but did not develop WRFPLUS. Barker et al. (AMS, 2005) originally developed WRFDA (for 3D-Var). These two latter references should be used as references for WRFDA. Barker et al. (2012) is an update on software development for WRFDA, and doesn't even mention WRFPLUS. The appropriate references need only be given at the first mention of these particular models, and do not need to be repeated throughout as references to the entire model. The exception is when discussing a particular aspect of those works.

15. Throughout, the author should use bold characters for vector notation (i.e., x, y, q, k_{co2}). These would be particularly illustrative on p. 10, lines 21, 24, and 31 to indicate whether the denominator or numerator of $\partial q_{co2}/\partial k_{co2}$ is a vector.

16. p. 5, line 7 and p. 14, line 23. You mention that L-BFGS-B can be used to calculate the posterior covariance, which is true although robustness of this approach with regards to the initial inverse Hessian estimate is an issue when using this algorithm, see Bousserez et al., QJRMS, 2015. As mentioned elsewhere in our review and this manuscript, Lanczos CG can also be used to estimate posterior error due to the eigen decomposition (well documented). Thus, the ability to calculate posterior error estimations is not a valid distinction between these two. Further, calculation of posterior error is not included for either method in this work. So while this could be mentioned in the introduction or discussion of future work, the methods section should only refer to methods that are actually used in

this work or ones that provide reasoning for why you used a particular approach.

17. p. 5, the term “cost function” is used 13 times on this page alone. It is suggested to reduce “cost function gradient” to “gradient”.

18. A comparison of Lanczos CG and LBFGS-B based solely on cost function reduction and RMSE is not sufficient. The authors should be more instructive and explicit as to the tradeoffs between them. In regards to p. 6, line 20, and p. 14, line 21: How much less memory does Lanczos CG require for your particular application, as a percentage? Is that a good reason for choosing it over L-BFGS-B in this case? Are there ways to reduce the memory requirements of each? The most accurate Lanczos CG algorithm requires storing all basis vectors and performing full reorthogonalization after each iteration. Do you include that step? If not, why? This is a salient topic, since you discuss the loss of conjugacy later in the manuscript. Also, how can Lanczos CG be adapted for parallel computation in a way that differs from L-BFGS-B? The name of an algorithm or a reference should be included. What are the respective wall-clock times of the two methods? Lastly, It’s also not clear why one is more amenable to parallel programming than the other (p. 14, line 21), as both are sequential techniques, unless that is strictly a consequence of the aforementioned memory requirements.

19. p. 6, line 20: Lanczos CG provides approximations of both the leading eigenvalues and eigenvectors (eigenmodes), not only the former.

20. p. 7, lines 8-9: Does VPRM calculate fluxes at the grid-scale in every time step? You can scale fluxes whether they are provided online or offline.

21. p. 7, line 15: So a tagging scheme for source specific CO2 has been implemented as well? This might present an interesting feature for testing the adjoint sensitivities and 4D-Var system, or performing low-dimensional analytic inversions.

22. p. 7, line 32: Well, to be more precise about the wording, the adjoint is just the backward sweep, although for nonlinear systems it would need information from the forward sweep. I'm not sure what that information would be though, for a linear CO2 simulation.

23. p. 10, line 29: The emission-normalized sensitivities can also be found by dividing the full sensitivities by the emissions. Two separate simulations are not required. You might say that you calculated them this way, but to suggest "this is done" in general by a specific approach is misleading.

24. p. 11, line 5: I realize this is more of a numerical demonstration than scientific result, but it is strange to define the adjoint forcing for tower observations to be at the surface rather than tower height, as in practice these types of measurements would have a greater emissions footprint (hence the rational for using a tower . . .).

25. Section 3.3: It seems like accuracy should be evaluated first, before presenting the sensitivity results in section 3.2.

26. Equation 8: What value used for $\triangle x$? It can sometimes be difficult to find a perturbation value that balances truncation and roundoff error when using this equation to verify adjoint sensitivities.

27. p. 12, line 23: By assuming $\mathbf{B}^{-1} = \mathbf{0}$ and $\mathbf{R} = \mathbf{I}$, the pseudo-data case ignores how uncertainties will affect the convergence of Lanczos-CG and LBFGS-B. How would the performance of these two approaches differ with imperfect observations? With an unbiased prior? Determining the correct treatment of $\mathbf{B}$ and $\mathbf{R}$ is an active research area, which the authors do not address. Do the authors plan to explore more realistic covariance definitions in the future? At a minimum, this should be discussed in Section 4.

28. p. 12, line 32-24: Did the authors confirm a loss of conjugacy in the Lanczos basis vectors? Also, did the authors make any attempt to force conjugacy through full re-orthonormalization (e.g., Modified Gram Schmidt)? That mechanism is built in to release version 3.6 of WRFDA. While re-orthonormalization uses extra memory, that resource requirement is often very small relative to that of the model integrations. The authors should justify a decision that adds iterations to the optimization. After including full re-orthonormalization, the number of iterations for Lanczos CG to converge in each outer iteration should be proportional to the degrees of freedom (DOF) constrained by the chosen observations (see, e.g., Rodgers, 2000), entirely independent of the conjugacy issue. At that point, the necessity of multiple outer iterations would be caused by a nonlinearity in the forward model, possibly the PBL treatment or convective transport. The authors make no attempt to characterize such a nonlinearity that would necessitate using a nonlinear optimization strategy.

29. Section 3.4: While interesting and valuable, numerically, there should be some statement with regards to the unphysical nature of the test setup, to emphasize that this is strictly a numerical test and not the expected level of performance (in terms of cost function reduction or RMSE) that would be obtained in a real inversion.

30. In Section 4, the third paragraph needs a topical sentence. Also, the two sentences “We evaluated ... sensitivity.” should be combined into one and made more concise.

31. p. 15, lines 1-11: While mentioning the ability to use different aggregation techniques may illuminate a budding area of research to the reader, the details given should be both accurate and concise. The authors’ discussion of smoothing and aggregation error (i.e., from Turner et al., 2015) are based in the assumption that no correlation is used in $\mathbf{B}$. Taken out of context, this could be very confusing to

the reader. Full non-ambiguous coverage of that topic would require more than a paragraph, but such a description is not appropriate for this section. Indeed, large portions of Section 4 (paragraphs 4, 5, and 6) ought to be rewritten or removed. Many of the references are out of date, and do not represent the state of the science.

32. p. 15, line 18-21: Other areas to improve upon would be more accurate treatment of data and model (i.e. transport and representational) errors in R, and error correlations in B, and posterior error estimation.

33. Section 5: I'm not sure this meets the requirements of GMD, and may delay the publication of this work until the code is publicly available.

Technical Corrections:

1. Add appropriate punctuation to Eqs. 1, 2, 3, 4, 6, and 7.

2. p. 1, line 21: Remove "inversion" at the end of the sentence, as it is implied in the first half of this statement.

3. p. 2, line 16: "LDMZ" should be changed to "LMDZ"

4. p. 2, line 17: Change "inverse" (noun) to "invert" (verb).

5. p. 3, line 3: "LPDM" is undefined. Possibly define and change "Lagangian particle backward trajectory model" to "Lagangian particle dispersion model (LPDM)" on p. 2, line 33.

6. p. 4, line 12: "Where" to "where"

7. The first term in parentheses in Eqs. 3, 4, and 6 need transpose operators. Additionally, it would be less confusing if brackets and braces are used in addition to parentheses where warranted.

8. p. 6, line 9: Correct “innoviation” to “innovation”
9. p. 6, line 11: Remove “Eq. 7”, since you are referencing the very next line of the text.
10. p. 6, line 20: Correct “lead” to “leading”
11. p. 7, lines 12, 16: “inner” to “inert”
12. p. 7, lines 14-16. Combine the two sentences that both state this category does not apply to CO2.
13. p. 7, line 20: “(Zhang et al.,” to “Zhang el al., (”
14. p. 7, line 30: Correct “simplied” to “simplified”
15. p. 8, lines 12 and 16: Correct “inner” to “inert”
16. p. 8, line 12: Correct “use” to “uses”
17. p. 9, line 1: Change “chemistry vertical mixing” to “vertical mixing of chemical species”
18. p. 9, line 3: “dynamical” to “the dynamical”
19. p. 9, line 24: “set up” to “setup”
20. p. 10, line 5: “simulation spans” to “simulations span”
21. p. 10, line 7: “condition” to “conditions”
22. p. 11, line 3: The reference should be to Gerbig et al. (2008). Also, use the correct parenthetical format for inline references.
23. p. 11, line 3: Change “footprint at a receptor” to “footprint of a receptor”

24. p. 11, lines 9, 11, 12, 15: The figure references are off by 1.

25. p. 11, line 18: Correct “no shown” to “not shown”

26. p. 12, line 23: Correct “identify” to “identity”

27. p. 12, line 28: Correct “facotr” to “factor”

28. p. 12, line 30-31: Lanczos-CG is repeated twice. Also, use either “Lanczos CG” or “Lanczos-CG” throughout the document.

29. p. 13, lines 2 and 21: cost function needs an article, such as “the”

30. p. 13, lines 2-4: Add commas before and after “by the 30th iteration”.

31. p. 13, lines 16-17: The opening to this sentence, “Starting at 2336.5 mol km-2 h-1,” is confusing or out of place.

32. p. 13, lines 24: change “the Lanczos” to “Lanczos” for consistency

33. p. 13, line 26: extra “the”

34. p. 14, line 8: Change “system” to “systems”

35. p. 14, line 10: Modify, “Such configuration”, which is grammatically incorrect.

36. p. 14, line 11: Change “incurring” to “requiring”

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-289, 2016.